LEVEL II

This research was sponsored by the
Defense Advanced Research Projects
Agency under ARPA Order No.: 3699
Contract No.: MDA903-79-C-0192,
Monitored by Dr. J. D. Fletcher

ARPA Order 3699

A STRUCTURAL APPROACH TO THE
VALIDATION OF HIERARCHICAL
TRAINING SEQUENCES

TASK 1 TECHNICAL REPORT

DTIC
ELECTE
APR 28 1980

A

79-12 21 017

John R. Bergan, Principal Investigator
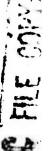Anthony A. Cancelli, Co-Principal
Olga Towstopiat, Research Associate

DISTRIBUTION STATEMENT A

Approved for public release
Distribution Unlimited

ADA083731

411726

Technical Report Number 1

Report Summary

This project is aimed at developing the technology necessary to conduct cost effective and efficient validations of the sequencing of instruction used in the training of military occupational specialties. The specific objective covered by this technical report was to validate task domains indicating how tasks are grouped into skill classes. A total of 317 subjects were tested on four algebra skill domains constructed from the Precision Measuring Equipment Curriculum of the Air Force Advanced Instructional System.
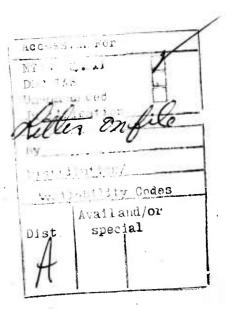
Latent structure techniques recently developed by Leo Goodman at the University of Chicago were used to validate the hypothesized domains. The first step in the analysis was to construct a set of models representing hypotheses about the tasks under examination. The models developed for use in the present analysis assumed three basic classes of individuals for tasks in an hypothesized domain. These classes included masters of the skill represented in the domain, non-masters, and individuals in transition between non-mastery and mastery. Non-masters were characterized as failing all items in the domain, and masters as passing all items. Transitional individuals were assumed to respond inconsistently in a manner congruent with the assumption that they were still in the process of acquiring the concept or rule underlying mastery of the tasks in the domains under examination. Models asserting that tasks were in the same domain were compared to models asserting that the tasks were unrelated.

A Texas Instrument 745 terminal purchased for the project was used in testing the extent to which the hypothesized models accurately represented the observed performance of the subjects. The analysis revealed three domains representing skill classes instead of the four hypothesized.

The identification of domains was essential to subsequent research planned for the project dealing with the ordering of domains. Clearly it would n be possible to order classes of skills in training sequences if evidence were lacking supporting the existence of skill classes. A major contribution of research described in this technical report was the discovery of clearly defined task domains.

A second important finding was the discovery that tasks within a domain may vary in difficulty level. This finding raises questions about generalization during the course of learning to master domain tasks. These questions may have far-reaching implications for training. More specifically, it may be possible to use information about difficulty level within a domain to determine where to begin instruction for the domain. This possibility has significant implications for training efficiency.

Narrative Table of Contents

Rationale for and Objectives

of the Proposed Research

Since the time that Robert Gagné (1962) introduced his learning-hierarchy model in the early 1960's, there has been a growing recognition of the usefulness of empirically validated hierarchical learning sequences in teacher based, computer assisted, and computer managed training programs aimed at promoting the acquisition of basic math and science skills or at the development of performance capabilities related to various technical specialties pursued in military and industrial settings (Glaser, 1976; Glaser & Nitko, 1971; Glaser & Resnick, 1972; Nitko & Hsu, 1974; Resnick, Wang & Kaplan, 1973; White, 1973, 1974). However, despite the recognized usefulness of hierarchies, validated hierarchical sequences that can be applied in training are lacking. Moreover, there is at present, no adequate, practical technology for conducting hierarchy validations. Unless such a technology is developed, the contribution that validated sequences could make to training will not be realized.

The validation of a learning hierarchy requires the testing of three hypotheses. One is that the specific trainee responses measured in the validation process represent response classes defining skills capable of being applied under a range of different stimulus conditions (Gagné, 1977). The second is that subordinate skills in a hierarchy are prerequisite or necessary to superordinate skills (Gagné, 1977), and the third is that prerequisite skills mediate transfer for superordinate skills (Gagné, 1977). The present project is designed to investigate research questions related to the testing of these hypotheses for the purpose of establishing guidelines that can be used in the development of a technology for hierarchy validation.

The Need for Validated Hierarchies

The need for validated hierarchies stems from their recognized potential value in training and from the fact that there are no adequately validated hierarchies in use in training programs today. Validated hierarchies could make two kinds of contributions in training. One of these relates to issues in instructional design, the other to assessment.

The Potential Role of Hierarchies in Instructional Design

The central advantage claimed for hierarchies in the area of instructional design has to do with the development of instructional sequences to facilitate transfer of learning. In numerous places in the literature, Gagné has advanced the view that lower level subordinate skills which are prerequisite to superordinate skills at higher levels in a hierarchy mediate transfer for the superordinate skills to which they are related (e.g., Gagné, 1962, 1968, 1973, 1977). The implication for instructional design is that instructional sequences should be arranged so that prerequisite skills are available to the trainee at the time that superordinate skills are to be mastered (Gagné, 1973).

Advocates of the learning-hierarchy view have pointed out that instructional sequences which ensure that prerequisite skills are available at the time of learning may produce highly beneficial results (e.g., Gagne, 1973; Glaser & Resnick, 1972). A sequence which takes into account prerequisite skills maximizes the likelihood that trainees will have appropriate prerequisite competencies at the time they are needed for superordinate-skill learning. On the other hand, a sequence developed without consideration for prerequisite relations leaves the question of whether or not trainees possess needed prerequisite competencies to chance. The result may be that some trainees will fail to master superordinate skills because they lack the prerequisites to superordinate skill mastery.

The Potential Role of Hierarchies in Assessment

The main advantage of empirically validated hierarchies with respect

to assessment relates to the problem of adapting instruction to the needs

of individual trainees. Given validated hierarchies, tests may be developed

to individualize the placement of trainees in an instructional sequence

(Glaser & Nitko, 1971; Nitko & Hsu, 1974; Resnick, Wang, & Kaplan, 1973).

Placement tests based on validated hierarchies may be used in the initial

phases of instruction to determine the point in an instructional sequence

which will enable a trainee to encounter readily attainable goals and at

the same time to avoid activities related to objectives that have already

been mastered. In addition, placement tests may be used at the end of a

sequence to determine what has been learned and thereby to establish what

should be taught next (Nitko & Hsu, 1974).

The Current Lack of Validated Hierarchies

White and Gagné (1974) have noted that although the learning-hierarchy

model has had some influence on the development of instructional materials

it has not yet had the wide application that might have been expected. One

apparent reason for the **fa**ilure of the learning-hierarchy model to have a

greater impact on training than it has had is that there are currently no

adequately validated hierarchies that could be used in training programs.

During the period since Gagné (1962) introduced the learning-hierarchy

model, there have been several studies attempting to validate isolated

hierarchical sequences (White & Gagné, 1974). However, early investigations

on hierarchies were marred by serious methodological flaws (White, 1973).

White (1973, 1974) suggested modifications in hierarchy validation procedures

which eventuated in marked improvements in validation techniques. Despite

these advances, adequate hierarchy validation has not yet been achieved.

As indicated in the initial paragraphs of the proposal, adequate hierarchy validation requires the examination of three hypotheses. Two of these three hypotheses have never been effectively tested in hierarchical research.

The hypothesis that skills in a hierarchy represent definable response classes has never been tested in hierarchy investigations. A few attempts have been made to assess the assumption that prerequisite skills mediate transfer for superordinate skills, but research in this area has had methodological flaws. Cotton, Gallagher, and Marshall (1977) have recently reviewed the literature on the transfer hypothesis and have concluded that Gange's transfer assumption has never been tested. Gagne's third hypothesis, the prerequisite-skills assumption, has recently been subjected to effective study (White, 1974). However, the validation procedures used to examine the prerequisite-skills assumption are extremely time consuming and may not be suitable for broad scale application.

Advances in Statistics that Make a Practical Technology for Hierarchy Validation Possible

A major reason for the lack of progress in hierarchy validation described above is that until recently appropriate statistical procedures have not been available to test hypotheses germain to the development of effective, practical procedures for validating hierarchies. A number of procedures have recently become available which should make it possible to conduct hierarchy validations in a practical and effective way.

New Techniques for Validating Prerequisite Relations. During recent years Gagne's prerequisite-skills assumption has served as a focal point for efforts to develop statistical procedures for use in hierarchy validation. White (1973) has shown that techniques used to assess prerequisite relations by Gagne and his colleagues in early hierarchy research were

inadequate in that they failed to provide a statistical test for prerequisite associations whi   ook into account errors in measurement. More recent research on prerequisite relations using a variety of scaling techniques including scalogram analysis (Guttman, 1944), multiple scalogram analysis (Lingoes, 1963), and the ordering theoretic method (Bart & Airasian, 1974; Bart & Krus, 1973) has been faulted on similar grounds. None of these procedures provides a suitable statistical test for prerequisite relations (Airasian, Madaus, & Woods, 1975; Dayton & Macready, 1976; White, 1974).

During recent years a number of attempts have been made to develop procedures to test Gagne's prerequisite-skills hypothesis statistically (Emrick & Adams, Note 2; Murray, Note 3; Proctor, 1970; White & Clark, 1973). Dayton and Macready (1976) have shown that each of these procedures represents a special case of a general latent-structure model which has the advantage of being capable of testing for prerequisite relations in both linear and nonlinear hierarchies. Goodman (1974, 1975) has also developed a latent-structure approach and a related model for scaling response patterns, both of which can be used to test for prerequisite associations in linear and nonlinear hierarchies.

New Techniques for Validating Positive Transfer. Although attempts to establish statistical techniques for use in hierarchy validation have focused mainly on Gagné's prerequisite-skills hypotheses, the need for procedures to examine Gagné's second major hypothesis, the positive-transfer assumptions are equally great. A recent review by Cotton, Gallagher, and Marshall (1977) attests to this fact. As indicated above, these investigators failed to find a single published study which provided a suitable test of Gagne's positive transfer assumption. Bergan (in press) has shown that structural equation models based on Sewall Wright's (1921, 1960) pioneering work in path analysis can be used to assess positive transfer in a learning hierarchy.

Structural-equation procedures based on regression analysis (Kerlinger &
Pedhazur, 1973) are available for use with interval scale dependent measures
(Duncan, 1975; Heise, 1975). In addition, Goodman (1972, 1973a, 1973b) has
developed structural-equation techniques involving the use of log-linear
models (Bishop, Fienberg, & Holland, 1975) that can be applied with dichoto-
mous and polytomous scores of the types typically used in hierarchy valida-
tion.

New Techniques for Domain Validation. As indicated above, Gagné (1977)
assumes that the skills in a learning hierarchy represent response classes
rather than discrete behavioral capabilities. For example, within the
learning-hierarchy viewpoint, it is assumed that a trainee who possesses a
skill such as multiplying two mixed numbers will be able to use that skill
to solve a broad range of similar problems.

One of the major problems in hierarchy validation is to determine
whether or not the items on a test of skill performance measure the trainee's
ability to perform the full range of behaviors included in the response
class assumed to be represented in the skill under examination. Hively,
Patterson, and Page (1968) used the term item domain to refer to the response
class associated with a given skill. In addition, Hively and his colleagues
developed a set of rules for generating test items falling within various
domains. Since the early work of Hively and his associates, other investi-
gators have elaborated on the concept of item domain and have attempted to
develop item generating procedures for various types of domains (Shoemaker,
1975).

Although awareness of the need to determine empirically the extent to
which specific test items represent an item domain has existed for some time,
statistical procedures for empirically validating item domains associated
with different skills have been lacking. For example, White (1974), in an
article on hierarchy validation, discussed the need for determining statis-

tically the extent to which different items assessed the same skill, but
was forced to conclude that there were no available statistical procedures
for making such a determination.

The Goodman (1975) response scaling technique and the Dayton and
Macready (1976) latent-structure model are both suitable for use in empiri-
cally validating an item domain. For instance, to test the hypothesis that
a set of items belong within the same domain using the Goodman scaling
technique, one would hypothesize a scaling model composed of two scale
types. One of these would represent those learners who had acquired the
skill being assessed by the items in the domain under investigation.
Trainees in this group would be expected to pass all domain items presented
to them. The second scale type would represent learners who had not acquired
the skill in question. Trainees in this group would be expected to fail all
daomin items which they encountered. Either the chi-square goodness-of-fit
or likelihood-ratio statistic can be used to test the fit of a model of this
type to a set of data collected on item performance in the domain targeted
for study.

A Structural Approach to Hierarchy Validation. The present research
combines use of the Goodman (1974) latent structure techniques with
structural equation procedures in which may be termed a structural approach
to hierarchy validation. The research examines the validity of item domains
in a hierarchy and addresses both Gagne's prerequisite-skills and positive-
transfer hypotheses as these asusmptions relate to the task of developing
practical procedures that can be applied in hierarchy validation in domain-
referenced assessment and training design. The hierarchical relations
selected for examination involve basic algebra skills included in military
training. The specific skills targeted for study have been selected
from the Precision Measuring Equipment Curriculum of the Advanced

Instructional System (AIS), an individualized training program operated by the Airforce at Lowrey Airforce Base. Analysis of these skills in the present project not only affords general guidelines for the validation of military training sequences, but also provides direct information that could be used to improve the efficiency and effectiveness of the precision measurement instructional unit.

Hierarchy Research Needs

Although adequate sta, .ical procedures for examining hierarchical relations are now available, information is lacking on how to go about the validation process. Three kinds of research needs must be met before it will be possible to determine the most efficacious procedures for validating hierarchical associations. One of these involves the issue of how skills should be measured in validating the prerequisite-skills hypothesis. The second has to do with skill measurement in validating the positive-transfer hypothesis, and the third deals with domain validation in hierarchical sequences.

Needs Related to Prerequisite-Skills Validation. One of the initial steps in hierarchy validation is to test for hypothesized prerequisite relations in the hierarchy under examination. Two strategies have been suggested for accomplishing this task. Research is needed to determine whether or not these two procedures yield different results.

One of the strategies used in prerequisite-skills validation is the psychometric approach (Resnick, 1973; Wang, 1973). In this approach, trainees are tested on skills under examination in a hierarchy, and a statistical procedure is applied to determine the existance of prerequisite dependencies. Some years ago White (1973) criticized the psychometric approach on the grounds that it does not control for random forgetting. White took the position that skills in a hierarchy may be forgotten in a different order than the order in which they are learned. In accordance with this position,

White (1974) argues that validation of the prerequisite-skills hypothesis requires a validation procedure in which learners who do not initially possess the skills in a hierarchy are taught the skills. He further suggested that testing for skill acquisition should be conducted during the course of learning rather than when instruction has been completed.

In support of the assumption of random forgetting, White cited only one study, an early investigation by Gagné and Bassler (1963). There are a number of reasons why the Gagné and Bassler study does not provide convincing evidence for the random forgetting assumption. First, adequate statistical procedures for testing the prerequisite skills hypothesis were unavailable at the time of the Gagné and Bassler investigation. Thus, it is not certain that all of the prerequisite relations that were assumed to be shown by the data actually did exist (White, 1976). Second, at the time of the investigation, there were no statistical techniques to assess the extent to which observed differences between learning and retention reflected measurement error as opposed to forgetting. Finally, the retention test which Gagné and Bassler used involved items which were different from the tiems used to assess learning. Thus, what Gagné and Bassler called a retention test could also be described as a test of generalization.

Recognition of the lack of convincing evidence provided by the Gagne and Bassler study has recently led White (1976) to suggest that the psycho-metric procedure ought to be reconsidered for use in hierarchy validation. The widespread application of hierarchical sequences in military training will require the validation of vast numbers of hierarchies. The psychometric approach to testing the prerequisite-skills hypothesis is much more efficient than the instructional strategy advocated by White. If it were possible to use the psychometric approach in the validation process and attain accurate results, a huge savings in time and personnel would be realized . In view of the superior efficiency of the psychometric approach and the lack of

convincing evidence contra-indicating the use of the approach, research to assess the efficacy of the psychometric technique is clearly warranted. In this regard, there is a need to determine the extent to which hierarchical models validated under White's instructional strategy match models validated psychometrically. The present project is designed to meet this research need.

As indicated in the discussion of the Gagné and Bassler study, the extent to which skills are retained in the order in which they are learned has implications with respect to the utility of the psychometric approach. Skill retention may be affected not only be forgetting processes, but also by the kinds of experiences the learner has after training has been completed. For example, the extent to which an individual uses skills on the job after a training progr m has been terminated may influence skill retention. In order to establish fully the utility of the psychometric validation strategy there is a need for additional research on the question of whether or not skills are forgotten in a different order than the order in which they are learned. Such research should include not only the examination of retention shortly after the completion of training, but also the study of retention in the post-training work environment. The present project addresses this research need.

Needs Related to Positive-Transfer Validation. As indicated above published studies assessing Gagné's positive-transfer hypothesis are lacking. One possible reason for this lack is that procedures advocated for testing positive transfer are difficult and time consuming to implement. Many investigators, particularly those studying complex hierarchies involving many connections have dealt with the issue of transfer by ignoring it and focusing instead on the validation of prerequisite relations (White & Gagné, 1974).

Validation of Gagné's positive-transfer hypothesis has generally been conceptualized within a transfer-of-training paradigm. White and Gagné (1974) suggest a validation strategy which illustrates this fact. The White and Gagné approach involves the following steps: First, choose as many prerequisite relations in the hierarchy under consideration as can be examined within existing constraints on time and resources. Second, for each connection to be studied, identify groups of learners who possess all relevant prerequisite skills, but who lack the specific prerequisite and superordinate skills targeted for study. Third, conduct a standard transfer-of-training experiment in which half of the learners receive training on the superordinate skill. Positive transfer is indicated if learners receiving prerequisite skill training perform significantly better on the superordinate-skill training task than learners who do not receive prerequisite skill instruction.

As indicated above, Bergan (in press) has shown that Gagné's positive-transfer hypothesis can be tested using structural equation models. Within a structural-equation approach, direct and indirect effects among a set of variables can be examined in the absence of an experiment involving random assignment of individuals to treatment conditions (Duncan, 1975; Goodman, 1972; Heise, 1975). For example, in the case of interval scale data, the direct effects of one variable on another can be assessed using ordinary least squares regression techniques (Duncan, 1975). The magnitude of the direct effect of the first variable on the second is given by a structural coefficient which in ordinary least squares regression analysis is the regression coefficient in the regression equation.

A structural approach to testing Gagné's positive-transfer hypothesis is potentially more efficient than the procedure suggested by White and Gagné. The increased efficiency derives from the fact that structural

equations can be used with the same data-collectionprocedures as those
employed in prerequisite-skills validation. Thus, for example, structural
equations can be used to examine positive transfer using White's (1974)
instructional procedure for prerequisite-skills validation. White's
instructional procedure requires less time and is more practical to implement
than the White and Gagne' (1974) transfer paradigm in that it necessitates only
one group of learners who are taught all skills in a linear sequence
whereas many groups learning different skills are needed to implement the
White and Gagne transfer procedure.

Structural equations can be used to achieve an even greater gain in
efficiency than that associated wtih the use of the White instructional
technique if they are coupled in positive-transfer validation with the
psychometric validation procedure. The psychometric procedure is, of course,
much more efficient than the White and Gagne approach in that all that is
required to implement the technique is to test a group of trainees.

To apply structural equations to test the assumption that prerequisite
skills mediate transfer for superordinate skills, prerequisite and superor-
dinate skills must first be identified. This can be accomplished using
prerequisite-skills validation procedures discussed above. After prerequisite
and superordinate skills have been determined, a structural model comprised
of equations expressing hypothesized effects of previously validated pre-
requisite skills on superordinate skills can be constructed. Data from
either the White instructional procedure of the psychometric procedure can
then be used in testing model-data fit.

It is possible that structural equations used either with White's
instructional technique or with the psychometric procedure would not yield
the same results as would be attained using the White and Gagne experimental
paradigm. If this were to occur, it could be argued that the White and

Gagné   approach provided a more valid demonstration of transfer

than a structural equation approach using prerequisite-skills validation

procedures in that the White and Gagné paradigm is experimental whereas the

structural-equation approach is not. However, if structural-equation pro-

cedures used with prerequisite-skills validation procedures could be assumed

to yield the same transfer relations as identified through the White and

Gagne paradigm, then a substantial gain in efficiency could be attained in

the validation process.

Research is needed to determine the extent to which structural-equation

techniques coupled with instructional or psychometric validation procedures

reveal the same transfer relations as those established through the use of

the White and Gagne experimental paradigm. The present project is designed

to meet this research need.

A corollary of the positive-transfer hypothesis that has appeared in

the literature from time to time (e.g., Cotton, Gallagher, & Marshall, 1977;

Resnick, 1973; Uprichard, 1970) is that not only will prerequisite skills

mediate transfer for superordinate skills, but also that instruction given

in the order suggested by the hierarchy will produce superior transfer to

that attained through the use of any other order. This hypothesis can be

investigated by teaching the skills under examination in all possible

orders (Cotton, Gallagher, & Marshall, 1977; Uprichard, 1970). The number

of connections that can be examined in this way is limited since the number

of possible orders becomes quite large when more than a few skills are

subjected to study. Thus, to validate hypothesized order effects in a large

hierarchy, it is necessary to conduct several studies on subsets of skills

in a manner analogous to the White and Gagne (1974) approach described above.

Cotton, Gallagher, and Marshall (1977) point out that the assumption

that hierarchical sequencing is maximally effective is important in deter-

mining the usefulness of hierarchies in designing instructional sequences.
They indicate further that the order assumption has never been adequately
tested. One of the aims of the present project is to test the hypothesis
that hierarchical sequencing produces optimal learning.

Needs Relating to Domain Validation. The validation of item domains
is an essential precursor to adequate examination of the other major
hypothesis involved in hierarchy validation. Without domain validation, it
is impossible to determine the extent to which test items reflect the
response classes that they are assumed to represent (Gagne, 1977). In the
absence of domain validation, failure to confirm either prerequisite-skills
or positive-transfer hypotheses could be attributed to the possibility that
the specific items used in validation did not adequately represent hypothesized
response classes for the skills under investigation. The focus of this
technical report is on domain validation.

Two questions must be answered in order to establish an adequate
technology for domain validation. One has to do with the size of domains.
An item domain may be either smaller or larger than hypothesized. For
example, two sets of items believed to assess separate skills in a hierarchy
could belong to the same domain (White, 1974). Adequate domain validation
requires that both the possibility of smaller than expected and larger than
expected domains be investigated. This technical report adresses the research
question.

The second question that requires attention with respect to domain
validation has to do with the comparability of domains validated psycho-
metrically and domains validated through the use of an instructional
paradigm. It is possible that item domains validated with an instructional
technique may differ from domains validated psychometrically. The amount
and type of skill training previously received by learners participating in

psychometric validation is not controlled and may differ in significant ways from the systematic training associated with instructional validation procedures. It is necessary to determine the extent to which item domains validated psychometrically match domains validated through instruction. Work to be completed during the second project year will address this research issue.

Project Objectives

Objectives for the present technical report focus on the attainment of Task 1 objectives. These include both outcome and enabling objectives.

Outcome Objective for Task 1. To validate psychometrically item domains for algebral tasks selected from an examination of the Precision Measuring Equipment Curriculum.

Enabling Objectives.

a. To task analyze algebra skills selected from the Precision Measuring Equipment Curriculum.

b. To construct and write item domains for each hypothesized domain.

c. Do construct of domain referenced test of items randomly selected from each domain.

d. To administer the test to approximately 100 subjects.

e. To score responses.

f. To construct and test latent class models to determine the extent to which hypothesized models fit (i.e., accurately represent) observed test performance.

Methods

Subjects

The subjects were 317 volunteers from a high school and university in the Southwest selected to represent a wide range of skill levels in solving algebra problems. Subjects ranged from high school freshmen taking a first course in basic mathematics to university students a number of whom had had college math courses. There were approximately equal numbers of males and females representing a broad spectrum of ethnic backgrounds. Approximately 88% were Anglo, 8% were Mexican American, and 4% were divided among blacks, native American Indians, and Asians. More subjects were used than the 100 originally intended for the study so that the full range of Algebra skills likely to be present in military trainees would be represented.

Task Analysis

In order to effectively develop hypotheses concerning the structure of the domains to be empirically validated, it was first necessary to task analyze the curriculum content for the purpose of identifying component skills. Task analysis procedures first emerged in military job training in the 1950's and 1960's as a systematic procedure for the identification of skills prerequisite to the performance of a designated task (Bernard, Note 1). Several methods of task analysis have since been proposed. One such procedure reported by Gagné (1968) involves the analysis of the intellectual skills making up a learning hierarchy. Learning is not conceived of by Gagné (1977) as the mastery of discrete behavioral capabilities but rather as the mastery of response classes mediated by intellectual skills. The most important of the intellectual skills is the "rule". If, as Gagné suggests, intellectual skills such as rules mediate mastery of homogeneous sets of items representing a single response class, then a task analysis of the

component skills making up a complex task will be an effective procedure in identifying domains which can be used as hypotheses for empirical scrutiny.

The curriculum content chosen for the present analysis is part of the Algebra programmed text in Block I of the Precision Measuring Equipment (PME) Curriculum in the Advanced Instructional System at Lowry Airforce Base. This content was decided upon with the Contracting Officer's Technical Representative. The task analysis involved the identification of component skills necessary for solving linear algebraic equations. As a result of the analysis, four component skills or domains were identified. They were: 1) term transposition, 2) expression transposition, 3) factoring, and 4) distributive property.

Term transposition refers to an intellectual skill involved in algebraic equation which requires the moving of single terms from one side of the equation to the other when attempting to isolate the variable for which one is solving. All tasks included in the present study required the subject to solve for x. A further analysis of each domain led to a specification of two dimensions within the domain each with two levels. They were: 1) the number of steps necessary to solve the equation (one step or two steps) and 2) the operations involved in the transposition (addition/subtraction or multiplication/division). Consequently, considering all possible combinations of these mutually exclusive dimensions, four tasks within this domain were hypothesized as important for analysis. An example of one of the tasks is $X + A = B$ where the subject is required to solve for X. This task would be term transposition involving one step and the addition/subtraction operation. While the hypothesis of initial concern implied that these dimensions would not significantly contribute to differences in subject performance, the dimensions were included to provide enough data for analysis in the event homogeneity among all items in the domain were not found.

Expression transposition refers to an intellectual skill involved in solving an algebraic equation which requires that the subject treat two or more terms in the equation as an expression. One dimension within this domain with three levels (one step, two steps or three steps) was further identified. An example of an expression transposition involving one step would be $X(N + R) = Y$.

Factoring is an intellectual skill which involves simplifying a polynomial by removing a common term before solving the equation. One dimension within this domain with four levels was identified (two, three, four and five steps). An example of a two step problem from the factoring domain is $NX + RX = Y$.

The fourth intellectual skill identified for domain structuring was distributive property. This involves multiplying an expression by a term in the course of solving the equation. Three steps, four steps and five steps were the three levels of the one dimension identified within the distributive property domain. $A(X + B) = D$ is an example of a three step task in the distributive property domain.

Domain Statements

Once the hypothesized domains were identified a domain-referenced test was constructed for administration to the subjects in the study. In order to construct the domain-referenced test it was first necessary to construct domain statements. The purpose of a domain statement is to more clearly specify those items to be considered as part of a domain. Traub (Note 4) identifies two types of domain statements. The first type, implicit domain statements are those which help specify the parameters of a domain. The specific items that can be used in constructing a test of the domain are only implicitly identified through the statement of these parameters. Domain specification procedures (Popham, 1978) and amplified objectives (Popham, 1974) are two types of implicit domain statements.

The second type of domain statement identified by Traub (Note 4) is the explicit domain statement. These involve procedures for the identification of all possible items making up the domain. Item form analysis (Hively, Maxwell, Rabell, Sension & Lundin, 1973) and facet analysis (Berk, 1978) are two types of explicit domain statement procedures.

One of the keys to determining the type of domain statement appropriate for a particular use is the structure of the curriculum content under analysis. More structured content allows for the use of explicit domain statements while implicit domain statements are more suitable for less structured content. Given the structured nature of the algebra content choosen for the present investigation, the item form analysis procedure (Hively et al., 1973) was used to identify the potential items available for inclusion in the domain-referenced test.

Figure 1 is an example of the item form which was constructed for the term transposition domain. The item forms for the reamining three hypothesized domains can be found in Appenlix A. As can be seen in the figure (page 20) the item form provides a detailed description of the characteristics of the items making up the domain and is composed of eight parts.

Title and General Description. In order to better understand an item form, one can first look at the Title and then the General Description. For example, in Figure 1 the title is "Solving Algebraic Problems Involving Term Transposition". The Title indicates that the task involves solving a problem by rearranging terms. The General Description indicates that three or four letters will be used to designate "unknowns" and that the subject will be asked to solve for one unknown; in this case, X. It can be solved by isolating that term on one side of the equation.

ITEM FORM 1

Solving Algebraic Problems Involving Term Transposition.

GENERAL DESCRIPTION

The individual is given an algebraic equation
involving three or four letters and asked to solve
for an unknown X, by isolating that unknown, on either
side of the equation.

STIMULUS AND RESPONSE CHARACTERISTICS

Constant for all Cells

Albegraic problems are presented in written form.
A written response is required.

Distinguishing Among Cells

Type of term transposition:
1) addition or subtraction (one step); 2) multiplication
or division (one step); 3) addition or subtraction (two steps);
4) multiplication or division (two steps).

Varying Within Cells

Type of problem presented and letters used to
represent variables.

CELL MATRIX   Set

Addition or subtraction (one step)	1
multiplication or division (one step)	2
addition or subtraction (two steps)	3
multiplication or division (two steps)	4

Figure 1. Item form 1 - Term Transposition

```
ITEM FORM SHELL


Materials

Pencil
Problem sheet with space for responses
```

Directions to Experimenter	Script
Place materials in front of individual	"Here is a sheet of problems. Solve for X in each problem. Try to do all of the problems. You will have as much time as you need.

SCORING SPECIFICATIONS

All items will be scored dichotomously (0 for incorrect response, 1 for correct response).

REPLACEMENT SCHEME

Use any of the problems from the replacement sets. Any letter of the alphabet may be substituted for another, except for X. Choose one from each of the following sets.

Replacement Sets

Set 1 :

$$X + A = B$$
$$X - A = B$$
$$A - X = B$$
$$A + X = B$$

Set 2:

$$X/A = B$$
$$AX = B$$

Set 3:

$$X + A - B = C$$
$$X + A + B = C$$
$$X - A + B = C$$
$$X - A - B = C$$

Set 4:

$$BX/A = C$$
$$ABX = C$$
$$X/A/B = C$$

Figure 1 (continued)

Item Form Cells. Item Form Cells represent the smallest classes of items into which a domain is stratified. It may include only one cell or many related cells. Cells are grouped into item forms on the basis of the dimensions identified. For example, in item form 1 in Figure 1 two dimensions, each with two levels or four cells, were isolated. These include 1 step addition/ subtraction, 1 step multiplication/ division, 2 step addition/ subtraction and 2 step multiplication division.

Item Form Shell. The Item Form Shell contains the common invariate components of all items generated by the item form. To produce an item, blank spaces in the shell are filled according to specifications given in the Replacement Scheme. The Item Form Shell includes the directions that are to be given to the subject by the examiner. The instructions must include the materials that are needed.

Item form 1 involves a relatively simple Item Form Shell. Since the problems are written, it is only necessary for the examiner to present the subject with a pencil and a list of problems, followed by little verbal instruction. The instructions are indicated in the script.

Replacement Scheme. The Replacement Scheme specifies how to choose values for each part of the variable parts of the item form. In item 1, the Replacement Scheme states that one item must be chosen from each of four sets. In addition, any letters other than X, may be substituted by other letters other than X. This allows for the construction of many items.

Replacement Sets. The replacement sets include items that were described in the cell matrix. For example, regarding term transposition, cell matrix 1 lists addition/ subtraction (one-step). Replacement set 1 lists four items that may be used in order to test term transposition for addition/ subtraction (one-step).

Stimulus and Response Characteristics. Stimulus and Response Characteristics are intended to describe and justify whatever behavioral analysis may underlie the properties or characteristics utilized in structuring the domain of items. These are grouped into three areas:

1) Characteristics that are constant for all cells in the item form. In item form 1, all of the problems are presented and answered in written form.

2) Characteristics that distinguish among cells. In item form 1, the distinguishing feature is the type of term transposition of concern.

3) Characteristics that are variable within cells. In item form one the distinguishing feature is the type of problem presented (addition vs. subtraction) and the letters used to represent unknowns in the problem.

Scoring Specifications. These describe the properties to be used to distinguish between correct and incorrect responses. In item form 1, a correct response is one which has appropriately isolated variable X. An incorrect response is one in which variable X has not been appropriately isolated.

Item Selection and Test Construction

Making use of the item forms, one item for each cell within a domain was chosen for the test. For purposes of analysis, each item was repeated twice with different letters in the equations (except for X which always remained the variable for which the subjects solved). Given the four cells and a total of fourteen dimensions, a test comprised of twenty-eight items was constructed. The items were randomly arranged on 215cm. by 27.9 mm. paper with sufficient space between items for the subjects to solve the equation. All equations required the subjects to solve for X. A copy of the test with correct responses included is included in Appendix B.

Scoring

Each pair of items representing a cell was scored $\underline{1}$, $\underline{2}$ or $\underline{3}$. A $\underline{1}$ indicates that neither of the two items was answered correctly. A $\underline{2}$ indicates that one of the two item pairs was answered correctly. If both items in the pair were answered correctly, a score of $\underline{3}$ was given.

Procedures

Testing was carried out in groups of about thirty. The participants were told that the purpose of the study was to determine how people solved algebra problems. After the test booklets were passed out, the experimenter gave instructions for responding to the test. Subjects were instructed to solve the algebra problems presented and to write their solutions in the test booklets provided. Subjects were instructed to attempt all problems and to provide solutions even in cases in which they were unsure of the answers. Following the instructions the subjects were told to begin the test and were assured that they would have as much time as necessary to complete the problems. During the course of the testing, the experimenter and an assistant monitored each subject's performance to insure that the task was udnerstood. The vast majority of the subjects comprehended what they were to do on the basis of the initial instruction. However, in one or two cases there were some questions. When this happened, the experimenter simply repeated the instructions for the individual having difficulty. In all cases the repeated instruction was sufficient to enable the individual to respond to the questions.

Models Used to Represent Domains

One of the reasons why research on rules in hierarchical sequences has been lacking is that until recently adequate statistical techniques have not been available to investigate equivalence and prerequisite relations in hierarchies. In recent years a number of techniques have been developed which make it possible to examine both prerequisite and equivalence relations in hierarchical sequences (Bergan, in press). As indicated above, the present investigation used latent class models (Goodman, 1974) to assess equivalence and ordered relations among algebra tasks.

Latent Class Models

Latent class models explain association in a contingency table in terms of a latent (i.e., unobserved) variable or set of latent variables each of which includes a set of latent classes. For example, in the present research latent class models were constructed to reflect variations in mastery for sets of items assumed to assess attainment of a set of item domains. The latent variable in this case was mastery variations. This variable included different latent classes, such as a mastery class and a non-mastery class.

A latent class model can be used to generate maximum likelihood estimates of expected cell frequencies which indicate expected response patterns under the assumption that the model being examined is true. The estimate for any particular response pattern is obtained by computing for each latent class the joint probability of the response pattern and the latent class. The joint probabilities are computed by an iterative process (Goodman, 1974). They are then summed across all latent classes and multiplied by sample size. For example, to determine the estimated expected frequency of passing all items in a given set, the joint probability of mastery and passing all items would be computed. Then the joint probability of passing all items and non-mastery

would be computed. This process would be continued until all latent classes had been included in the computations. Then the joint probabilities would be summed to give the probability of passing all items estimated under the model. The estimated expected cell frequency would be determined by multiplying this estimated probability of passing all items by sample size.

Various kinds of restrictions can be imposed on latent class models. For example, the assumption of a homogeneous item domain implies equivalence among tasks assessing domain mastery. This indicates that certain classes of trainees ought to perform in the same way across tasks. Latent class models can be used to reflect this kind of assumption. The probability of task mastery can be restricted to be equal across tasks. Moreover, probabilities can be set at a particular value such as 1 or 0. For instance, it may be assumed that the probabilitiy of performing a task correctly is 1 for a master and 0 for a non-master.

Testing Latent Class Models

Latent class models are tested by assessing the correspondence between observed cell frequencies and estimates of expected cell frequencies using the chi-squared statistic. When the correspondence between observed and expected frequencies is close, the value of X^2 will be low and the model being tested can be said to provide an adequate fit for the data. Clifford Clogg (Note 2) has developed a computer program which carries out the iterative process used to generate maximum likelihood estimates of expected cell frequencies and which computes the X^2 value to test the fit of a model to a data set. Clogg's program was used in the present investigation. Because it was necessary to cross-classify many different item sets during data analysis, Clogg's program was linked to another computer program called ALT. This program, which is described in Appendix D, was used to construct contingency tables from subjects' responses to test items. These tables provided the

input data for Clogg's latent structure program.

Model Characteristics

The latent class models designed for the present project were intended to distinguish between ordered and equivalence relations among algebra tasks. Although only those models reflecting equivalence relations were used in the present report, the models contained provisions which made it possible to distinguish between ordered and equivalence relations. Thus, to understand why the models were designed as they were, it is necessary to understand model distinctions involving the ordering and equivalence of tasks.

Consider a two-way table like the one below, cross-classifying performance on two algebra tasks, A and B. Each task is assessed in terms of performance on two items. Thus, a subject's score for each task may fall into one of three categories. zero right, one right, or two right. We shall designate these three categories by the numbers 1, 2, and 3 respectively.

Task B

	1	2	3
Task A 1			
2			
3			

In a table of this kind, a score of 1 on each task would suggest non-mastery. This response pattern would be reflected in the 11 cell in the table. A score of 3 on each task would suggest mastery. This pattern is reflected in the 33 cell. A score of 3 on task A and 1 on task B would indicate mastery of task A without evidence of mastery of task B. Scores of 2 would reveal inconsistent performance. Since the items for each task are identical, scores of 2 should

reflect errors which ought to occur at a relatively low frequency.

Given an ordered relation between tasks A and B, the number of responses in the 31 cell should be significantly greater than the number in the 32 cell. Under the assumption of ordering, a build-up would be expected in the 31 cell indicating that a significant number of subjects had mastered A without having mastered B. The 32 cell would be expected to have relatively few responses because the 2 category represents response errors for task B.

By contrast, if the tasks were equivalent, the 31 cell would not be expected to contain a large number of individuals. The 31 category, like the 32 category, would represent error responses. Under the equivalence assumption one of two relations between the 31 and 32 cells might be expected. Either the cells would be equiprobable or there might be a small, but nonetheless significantly greater number of individuals in the 32 cell than in the 31 cell.

As this discussion shows, a critical issue in determining whether two tasks form an ordered or equivalence relation is that of determining whether the hypothesis that the occurrence of responses in the 31 and 32 cells is equiprobable is supported by the data. If this hypothesis is rejected, it is necessary to determine whether the probability of a response in the 31 cell is greater than the probability of a response in the 32 cell for masters of task A.

Many of the equivalence models described in the present report include restrictions reflecting the equiprobability hypothesis. These restrictions are included to serve as a basis for distinguishing between ordered and equivalence relations.

In addition to distinguishing between ordered and equivalence relations, the models designed for use in the research differentiated between a variety

of types of equivalence relations reflecting homogeneous item domains. Inves-
tigators concerned with the problem of developing models to describe homo-
geneous item domains have taken the position that one criterion that can be
used in defining such domains is that at least some groups of individuals
should tend to respond to all items in the domain in the same way (Dayton &
Macready, 1976). For example, it is generally assumed that masters of an
item domain will pass all the items in the domain and that non-masters will
tend to fail all items in the domain (Dayton & Macready, 1976).

Varying assumptions have been made about the number of mastery classes
useful in defining homogeneous item domains. The models constructed for
task 1 activities hypothesized 4 classes of individuals: masters, partial
masters, non-masters, and individuals in transition between non-mastery and
mastery. It was assumed that masters would pass all domain items, that non-
masters would fail all domain items, and that partial masters would consistently
perform one out of two task problems correctly. Individuals in transition
between non-mastery and mastery were assumed to perform some items correctly
but not others.

Within the transitional category, the probability of passing any given
item was assumed to be less than 1. The fact that the probability of a given
level of performance could be less than perfect raises questions about item
difficulty within the transitional latent class. Varying assumptions can be
made about item difficulty. More specifically, it can be assumed that item
difficulty varies within the transitional latent class, or that it is equal
across items within that class.

The distinction between varying and equal item difficulty suggests the
identification of two forms of equivalence within a homogeneous item domain.
One of these may be called asymmetrical equivalence. Items which are
asymmetrically equivalent are items which are equivalent for masters, partial

masters or non-masters but which differ in difficulty level for transitional individuals. The second type of equivalence may be described as symmetrical equivalence. Items are symmetrically equivalent if they are of equal difficulty level in all latent classes.

Descriptions of Domain Testing Models

The D.A.R.P.A. Hierarchy Research Project was designed to make use of 8 latent class models which distinguish between ordered relations and asymmetrical and symmetrical equivalence relations. All but one of these models were used in the activities described in the present technical report. One of the models was designed to reflect ordered relations among tasks. This model will be used in the validation of prerequisite relations, which is the subject of the next technical report. One of the models affords a standard against which to compare equivalence models. The other 6 models reflect various forms of equivalence relations used in studying homogeneous item domains. The models used in domain validation are described below and are displayed visually on the next page. The E's and curved lines in the visual display indicate cells constrained to be equiprobable under a given model. The I's indicate cells for which the assumption is made that the probability of a given level response on task A is independent of the probability of any particular response level on task B. The X's indicate response patterns associated with specific latent classes. For example, the X in the 11 cell of H_1 indicates the association of the 11 response pattern with the non-mastery latent class.

The Independence-Equiprobability Model. The first model, designated H_0, asserts independence between task pairs and equiprobability between categories 1 and 2 for the task assumed to be the least difficult in the task pair. This model served as a standard against which to compare the other models tested.

Models Used in Establishing Item Domains

H_0

B

	1	2	3
A 1	E	E	I
A 2	E	E	I
A 3	E	E	I

H_1

B

	1	2	3
A 1	X	E	E
A 2	E	X	E
A 3	E	E	X

H_2

B

	1	2	3
A 1	X	I	I
A 2	E	E	I
A 3	E	E	X

H_2'

B

	1	2	3
A 1	X	I	I
A 2	I	I	I
A 3	I	I	X

H_3

B

	1	2	3
1	X	I	I
2	I	X	I
3	E	E	X

H_3'

B

	1	2	3
1	X	I	I
2	I	X	I
3	I	I	X

H_4

B

	1	2	3
1	X	I	I
2	X	I	I
3	E	E	X

1. The E's connected by curved lines indicate cells constrained to be
   be equiprobable. The I's indicate cells for which the hypothesis of
   independence prevails. The X's indicate cells reflecting response
   patterns associated with specific latent classes.

The equiprobability provision was included to make the model congruent with
other models being examined. As mentioned earlier, the central criterion
for distinguishing between ordered and equivalence relations is one asserting
equiprobability between certain task categories. The equiprobability provision
was included in model H_0, as well as some of the other models examined, to
provide a basis for distinguishing between ordered and equivalence relations.
If there had been any instances in which model H_0 provided an adequate des-
cription of tasks in the domain under examination, the hypothesis that the
tasks were not related would have been supported.

The Model of Symmetry. Model H_1 asserted symmetrical equivalence between
tasks. Model H_1 included 6 latent classes: a non-mastery class, a partial
mastery class, a mastery class, and 3 transition classes reflecting symmetri-
cal inaccuracies in responding. The 3 classes assuming inaccurate responding
each asserted equiprobability for one pair of cells in the table cross-classi-
fying the tasks under examination. For example, one of these classes asserted
that the probability of the 12 cell would be equal to the probability of the
21 cell. The second asserted that the probability of the 13 cell would be
equal to the probability of the 31 cell, and the third assumed that the prob-
ability of the 23 cell would be equal to the probability of the 32 cell. Be-
cause of the symmetrical nature of its equiprobability restrictions, this
model has been described in the literature as the model of symmetry (Bishop,
Fienberg, & Holland, 1975). The model of symmetry implies equal item diffi-
culty for the tasks under examination. Tasks for which this model provided
an adequate fit for the data were described as being symmetrically equivalent.

Asymmetrical Equivalence Models. Model H_2 included 3 latent classes, a
mastery class, a non-mastery class, and a class composed of transitional
individuals. Model H_2 assumed that masters would respond correctly to all
problems presented to them. Thus, in the mastery class the probability of

the 33 response pattern was restricted to be 1. Similarly, the model assumed that non-masters would fail all problems. Thus, in the non-mastery class the probability of the 11 category was restricted to be 1. It was presumed that in the unscalable category, the probability of a particular level of performance on one task would be dependent of a given level of performance on the other tasks, and that the 1 and 2 categories would be equiprobable for one of the tasks. The equiprobability restriction was included as a criterion for distinguishing between equivalence and ordered relations for reasons already discussed.

Model H_2' is a special case of model H_2. It is like model H_2 in all respects except that it does not include the equiprobability restriction imposed under H_2. Model H_2' was included to reflect the fact that two tasks may be equivalent even though the 1 and 2 categories of the more difficult task are not equiprobable. It may happen that the probability of a response in the 32 cell is greater than the probability of a response in the 31 cell. This is exactly the opposite of what is to be expected under the hypothesis of an ordered relation between tasks. When the hypothesis of equiprobability is rejected, but the probability of the 32 cell is greater than the probability of the 31 cell, it is appropriate to test models which assert equivalence, but which do not include equiprobability restrictions. Model H_2' is one such model.

Model H_3 included 4 latent classes, a non-mastery class, a partial mastery class, a mastery class, and a transitional mastery class. The partial mastery class was similar to the transitional class in that both reflected less than completely accurate responding on the part of examinees. However, model H_3 asserted that individuals in the partial mastery class consistently performed 1 out of 2 problems correctly on both tasks under examination for a given task pair. More specifically, the partial mastery class asserted

that for members of that class the probability of getting 1 out of 2 items correct for both tasks would be 1. The transition class did not assume this kind of consistency in partially accurate responding.

Model H_3' assumed four latent classes, a non-mastery class, a partial mastery class, a mastery class, and an unscalable class. The restrictions for non-mastery, partial mastery, and mastery classes were the same as those given for H_3. Moreover, similar restrictions were imposed for partial mastery.

Model H_3' differed from H_3 because it did not impose an equiprobability restriction in the unscalable category. The concept of partial mastery implies a significant number of individuals who get 1 problem right. Given this state of affairs, not only should a build-up of individuals in the 22 category be expected, but also it would not be unreasonable for the probability of occurrence of the 32 category to be greater than the probability for the 31 category. Model H_3' reflects the fact that equiprobability need not always occur in a model asserting equivalence between tasks.

Model H_4 is very similar to H_2. The difference between the two is related to the equiprobability restriction in the unscalable class. In asserting both independence and equiprobability, model H_4 necessarily makes the 21 and 22 cells as well as the 31 and 32 cells equiprobable in the unscalable latent class. Equiprobability does not obtain for the 11 and 12 cells because the 11 cell represents a separate latent class, i.e., the non-mastery class. Model H_4 restricts equiprobability in the unscalable class to the 31 and 32 cells. This is accomplished by making the 21 cell represent a separate latent class. The probability of the 21 response pattern in this class is restricted to be 1. The effect of this is to make the observed and expected cell frequencies for the 21 pattern equal. Thus the pattern contributes nothing to the value of X^2. With the exception of the restriction on the 21 cell, model H_4 is exactly the same as H_2. Like H_4, it contains mastery, non-mastery and transi-

tional latent classes. Moreover, the restrictions on the mastery and non-mastery classes ard the same as those for H_2. The unscalable category assumes independence between tasks with the 21 pattern ruled out of consideration. In addition, it asserts equiprobability for the 31 and 32 cells.

Results

Results of the model testing revealed three domains instead of the four hypothesized. Tables 1, 2, and 3 in Appendix C present the observed responses for the cross-classification of every possible task pair for each of the 3 domains. Table 1 shows the cross-classification for the term trasposition domain, Table 2 for the Expressions Transposition, and Table 3 for the newly formed Distributive Property-Factoring domain. In Table 1 the letters indicate the addition-subtraction (A) and multiplication-division (M) dimensions. In Table 2 the letters designate the expressions domain, and in Table 3 they stand for factoring (F) and distributive property (D) problems. Numbers in all three tables represent the number of steps required for problem solution.

The response patterns in the tables indicate various combinations of the number of correct responses for each task pair examined. For example, the 11 pattern indicates no correct responses on either task while the 33 pattern represents 2 correct responses for each task. Note the large number of responses falling in the 11 and 33 categories in the tables. These patterns represent the critical cells for establishing equivalence relations. Notice further that most task sets have about the same number of individuals in the 31 and 32 cells. The 31 cell represents individuals who have mastered one task, but have not begun to acquire the second task. As already indicated, given prerequisite relation between tasks, the number of individuals in the 31 cell would be expected to be larger than the number of individuals in the 32 cell. On the other hand, given an equivalence relation between the tasks, the number of individuals in both the 31 and 32 cells would be expected to

be small.

Tables 4, 5, and 6 in Appendix C present the results of model testing for the hypothesized domains. In the model testing process, all possible pairs of tasks within a given domain were compared. In addition, all possible comparisons were made between domains hypothesized to be adjacent.

Table 4 shows the chi-squared tests for all possible task pairs in the term transposition domain. The letters designating tasks refer to the addition-subtraction (A) and multiplication-division (M) dimensions for this domain. The numbers refer to the number of steps required for problem solution. For example, 1 refers to a problem requiring only one step for solution. The model testing process required the selection of a preferred model based on statistical comparisons among various models examined. To illustrate the comparison process, consider the results for H_0 and H_2 for the Al-Ml task pair given in Table 4. The X^2 value for model H_0 is 200.65 with 5 degrees of freedom, which is significant well beyond the .001 level. The X^2 value for model H_2 is 1.18 with 3 degrees of freedom which has a p value of about .90. Model H_0 and H_2 are hierarchical. That is, H_2 contains all of the characteristics of H_0 plus 2 additional characteristics. These additional characteristics reflect the inclusion of a mastery and non-mastery latent class under H_2. Model H_2 has 3 degrees of freedom, whereas H_0 has 5. The loss of 2 degrees of freedom reflects the inclusion of the non-mastery and mastery latent classes. Because H_0 and H_2 are hierarchical, they can be compared statistically. The X^2 for H_2 can be subtracted from the X^2 for H_0. The result will be a X^2 with 2 degrees of freedom. In the case of the Al-Ml task pair, the subtraction of H_2 from H_0 yields an X^2 of 198.47 with 2 degrees of freedom, which is significant far beyond the .01 level. Thus, model H_2 provides an excellent fit for the data. Moreover, none of the models improve over H_2. Consequently, H_2 was selected as the preferred model for the Al-Ml

task pair. Not all of the models in Table 4 are hierarchically related.
For example, H_1, the symmetry model, is not hierarchically related to either
H_0 or H_2. Consequently, it is not possible to compare H_1 direct with H_0 or
H_2.

Where possible, hierarchical comparisons were used as a basis for select-
ing preferred models. In those few instances where hierarchical comparisons
could not be made, indirect inferences involving assessments of goodness-of-
fit were used for selecting a preferred model. Preferred models are indi-
cated in Table 4 by asterisks.

The results on Table 4 show that in no case did model H_0 or H_1 provide
an acceptable fit for the data. Consequently, the hypothesis that the task
pairs under examination were unrelated and the hypothesis that they were
symmetrically equivalent could be rejected for all of the tasks investigated.

In all cases except one, one of the asymmetrical equivalence models pro-
vided an acceptable fit for the data. In some instances, the model including
an equiprobability restriction provided an adequate fit. In other cases, for
example in the case of task pair A1-A2, the equiprobability assumption was
rejected. However, the probability of being in the 32 cell was found to be
higher than the probability of being in the 31 cell. Consequently, it could
safely be concluded that the tasks for this pair were not ordered.

The one instance in which the hypothesis of equivalence relations was
rejected was that involving the A1-M2 task pair. The two tasks involved in
this comparison represented extremes in difficulty level within the item do-
main. The one-step addition problem was the simplest task in the domain,
whereas the two-step multiplication problem was the most difficult task.
These two tasks formed an ordered relation. As already indicated, the inves-
tigation of ordered relations will be the major topic of concern in the next

technical report. However, it may be pointed out at present that the ordered relation for the A1-M2 task pair signifies permeability in domain boundaries. Tasks A1 and A2 are in the same domain. A2 and M2 are not in the same domain. The fact that A1 and M2 are found to be in separate domains suggests that the boundaries between domains are not rigid. This finding, as will be discussed in the next technical report, may have far-reaching implications for the conceptualization of structural relations within and among item domains and therefore for training.

The results for the term transposition domain reflect a highly consistent pattern. As already indicated, the hypothesis of asymmetrical equivalence was supported in every instance except one. The asymmetrical equivalence observed in the domain reveals a highly structured arrangement of tasks. The tasks in the multiplication-division dimension are more difficult than those in the addition-subtraction dimension. Moreover, tasks requiring two steps for problem solution are more difficult than those requiring only a single step.

Table 5 presents the results of model testing for the expressions transposition domain. The letters designating tasks refer to the fact that the tasks are in the expression category, and the numbers indicate the number of steps required for problem solution. There were only 3 possible comparisons for this domain. In two of the three comparisons an asymmetrical equivalence model provided an acceptable fit for the data. In the third case, none of the models tested afforded an adequate fit. The reason for this is that for the E2-E3 task pair a greater number of individuals than expected under the models tested mastered the more difficult task before they had acquired the less difficult task.

Table 6 shows the results of model testing for the combined distributive property-factoring domain. The letters in Table 6 refer to factoring problems (F) and distributive property problems (D). As for the other tables,

numbers indicate the number of steps required for problem solution.

As in the case of the other two domains, the results for the combined distributive property-factoring domain reveal a highly consistent pattern. In most instances, one of the asymmetrical equivalence models provides a suitable fit for the data. However, in some cases, the model of symmetry fit the data to an acceptable degree. This suggests that at the higher levels of algebra skill, problems are more likely to be equivalent for all groups of individuals, including those in transition. This is understandable since those in transition with respect to higher level skills bring a broad background of subordinate skills to the task of solving higher level factoring and distributive property problems.

In only one case did a task not form an equivalence relation with other tasks. This was the case for the most difficult factoring task. Model testing revealed that this task was superordinate to all of the other distributive property and factoring tasks. Analysis of the characteristics of this task revealed that it required not only factoring, but also application of the multiplication operations used for the distributive property problems. An item form for this newly constructed domain is found in Appendix A (Item form 5).

Discussion

The results revealed in this initial technical report with respect to domain validation suggest that the mathematical skills examined form highly structured and extremely consistent equivalence relations. The empirically validated equivalence relations are summarized in Table 7 in Appendix C. The consistency of the findings provides a strong foundation for proceeding with the next phases of research planned for the project. The next task to be accomplished in the project involves the hierarchical ordering of domains. It would not have been possible to accomplish this task if consistent domains had not been revealed.

The differentiation between asymmetrical and symmetrical equivalence

relations requires comment. This differentiation was not anticipated in the original proposal. The possibility of the two forms of equivalence became apparent during model construction. The empirical validation of the two forms of equivalence that emerged through model testing is an important finding which raises questions that should enrich the value of research planned for the latter phases of the project. For example, the discovery of two forms of equivalence raises questions about within domain generalization. For example, it might be assumed that in the case of symmetrically equivalent tasks, the learning of either skill would imply generalization to the other skill. On the other hand, in the case of asymmetrical equivalence, it might be hypothesized that learning the more difficult will imply generalization to the less difficult skill whereas learning the less difficult skill may not ensure generalization to the more difficult skill. This possibility will be examined in research planned for the third project year.

It is important to note that basically two forms of asymmetrical equivalence emerged from the data analysis. In many cases, a model which did not contain a partial mastery class fit the data. Models in which this was the case tended to support the equiprobability hypothesis for scoring categories 1 and 2 (i.e., none right and one right) of the more difficult task. In some cases asymmetrical equivalence involved partial mastery. When this was the case, the equiprobability assumption tended to be rejected. What this suggests is that the greater the degree of inconsistent responding (i.e., 1 out of 2 problems correct), the greater will be the likelihood of a partial mastery class being required to achieve model-data fit.

Reference Notes

1. Bernard, M.E. Task analysis in instructional program development. Theoretical Paper No. 52. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, Wisconsin, May, 1975.

2. Emrick, J.A. & Adams, E.N. An evaluation model for individualized instruction (Research Report RC 2674) Yorktown Hts., New York: IBM Thomas J. Watson Research Center, 1969.

3. Murray, J.R. Statistical models for qualitative data with classification errors. Unpublished doctoral dissertation, University of Chicago, 1971.

4. Traub, R.E. Stirring muddy water: Another perspective on criterion-referenced measurement. Ontario Institute for Studies in Education, mimeo, 1975.

References

Airasian, P.W., Madaus, G.F., & Woods, E.M. Scaling attitude items: A comparison of scalogram analysis and ordering theory. Educational and Psychological Measurement, 1975, 35, 809-819.

Bart, W.M., & Airasian, P.W. Determination of the ordering among seven Piagetian tasks by an ordering-theoretic method. Journal of Educational Psychology, 1974, 66, 277-284.

Bart, W.M., Krus, D.J. An ordering-theoretic method to determine hierarchies among items. Educational and Psychological Measurement, 1973, 33, 291-300.

Bergan, J.R. The structural analysis of behavior: An alternative to the learning hierarchy model. Review of Educational Research, in press.

Berk, R.A. The application of structural facet theory to achievement test construction. Educational Research Quarterly, 1978, 3, 62-72.

Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. Discrete multivariate analysis. Cambridge: MIT Press, 1975.

Cotton, J.W., Gallagher, J.P., & Marshall, S.P. The identification and decomposition of hierarchical tasks. American Educational Research Journal, 1977, 14, 189-212.

Dayton, C.M., & Macready, G.B. A probabilistic model for validation of behavioral hierarchies. Psychometrika, 1976, 41, 189-204.

Duncan, O.D. Introduction to structural equation models. New York: Academic Press, 1975.

Gagné, R.M. The acquisition of knowledge. Psychological Review, 1962, 69, 355-365.

Gagné, R.M. Contributions of learning to human development. Psychological Review, 1968, 75, 177-191.

Gagné, R.M. Learning and instructional sequence. In F. Kerlinger (Ed.), Review of Research in Education (Vol. 1). Itasca, Ill.: Peacock, 1973.

Gagné, R.M. The conditions of learning (3rd ed.). New York: Holt, Rinehart & Winston, 1977.

Gagné, R.M., Bassler, O.C. Study of retention of some topics of elementary nonmetric geometry. Journal of Educational Psychology, 1963, 54, 123-131.

Glaser, R. Components of a psychology of instruction: Toward a science of design. Review of Educational Research, 1976, 46, 1-24.

Glaser, R., & Nitko, A.J. Measurement in learning and instruction. In R.L. Thorndike (Ed.), Educational Measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

Glaser, R., & Resnick, L.B. Instructional psychology. Annual Review of Psychology, 1972, 23, 207-276.

Goodman, L.A. A general model for the analysis of surveys. American Journal of Sociology, 1972, 77, 1035-1085.

Goodman, L.A. Causal analysis of data from panel studies and other kinds of surveys. American Journal of Sociology, 1973, 78, 1135-1191. (a)

Goodman, L.A. The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach. Biometrika, 1973, 60, 179-192. (b)

Goodman, L.A. The analysis of systems of quantitative variables when some of the variables are unobservable. Part I--A modified latent structure approach. American Journal of Sociology, 1974, 79, 1179-1259.

Goodman, L.A. A new model for scaling response patterns: An application of the quasi-independence concept. Journal of the American Statistical Association, 1975, 70, 755-768.

Guttman, L. A basis for scaling qualitative data. American Sociological Review, 1944, 9, 139-150.

Heiss, D.R. Causal analysis. New York: Wiley, 1975.

Hively, W., Maxwell, G., Rabehl, G., Sension, D., & Lunden, S. Domain referenced curriculum evaluation: A technical handbook and a case study from the Minnemast Project (CSE Monograph Series in Evaluation No. 1). Los Angeles: University of California at Los Angeles, Center for the Study of Evaluation, 1973.

Hively, W., Patterson, H.L., & Page, S.H. A "universe-defined" system of arithmetic achievement tests. Journal of Educational Measurement, 1968, 5, 275-290.

Kerlinger, F.N., & Pedhazur, E.J. Multiple regression in behavioral research. New York: Holt, Rinehart & Winston, 1973.

Nitko, A.J. & Hsu, T. Using domain-referenced tests for student placement, diagnosis and attainment in a system of adaptive, individualized instruction. In W. Hively (Ed.), Domain-referenced testing. Englewood Cliffs, New Jersey: Educational Technology Publications, 1974.

Popham, W.J. Criterion-referenced measurement. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Popham, W.J. An approaching peril: Cloud-referenced tests. Phi Delta Kappan, 1974, 56, 614-615.

Proctor, C.H. A probabilistic formulation and statistical analysis for Guttman scaling. Psychometrika, 1970, 35, 73-78.

Resnick, L.B. (Ed.). Hierarchies in children's learning: A symposium. International Science, 1973, 2, 311-362.

Resnick, L.B., Wang, M.C., & Kaplan, J. Task analysis in curriculum design: A hierarchically sequenced introductory mathematics curriculum. Journal of Applied Behavioral Analysis, 1973, 7, 679-710.

Shoemaker, D.M. Toward a framework for achievement testing. Review of Educational Research, 1975, 45, 127-147.

Uprichard, A.E. The effects of sequence in the acquisition of three set relations: An experiment with preschoolers. Arithmetic Teacher, 1970, 17, 597-604.

Wang, M.C. Psychometric studies in the validation of an early learning curriculum. Child Development, 1973, 44, 54-60.

White, R.T. Learning hierarchies. Review of Educational Research, 1973, 43, 361-375.

White, R.T. The validation of a learning hierarchy. American Educational Research Journal, 1974, 11, 121-136.

White, R.T. Effects of guidance, sequence, and attribute-treatment interactions on learning, retention, and transfer of hierarchically ordered skills. Instructional Science, 1976, 5, 133-152.

White, R.T., & Clark, R.M. A test of inclusion which allows for errors of measurement. Psychometrika, 1973, 38, 77-86.

White, R.E., & Gagne, R.M. Past and future research on learning
hierarchies. Educational Psychologist, 1974, 11, 19-28.

Wright, S. Correlation and causation. Journal of Agricultural
Research, 1921, 20, 557-585.

Wright, S. Path coefficients and path regressions: Alternative or
complimentary concepts? Biometrics, 1960, 16, 189-202.

Appendix A

Item Forms

ITEM FORM 2

Solving Algebraic Problems involving Expression Transposition

GENERAL DESCRIPTION

The individual is given a algebraic equation invol ing four or five
letters and is asked to solve for an unknown, X. This can be done by
isolating the unknown on one side of the equation. However, it can only
be done by moving an expression (set of variables) to one side of the
equation.

STIMULUS AND RESPONSE CHARACTERISTICS

Constant for all Cells

Algebraic problems are presented in written form. A written response
is required.

Distinguishing Among Cells

Type of expression:
1) transposition (one step); 2) transposition (two steps); 3) transposition
(three steps).

Varying Within Cells

Letters used to represent variables.

CELL MATRIX

one step	1
two step	2
three step	3

figure 2. Item form 2 - Expression Transposition

figure 2 (continued)

ITEM FORM SHELL

Materials

Pencil
Problem sheet with space for responses

Directions to Experimenter	Script
Place materials in front of individual	"Here is a sheet of problems. Solve for X in each problem. Try to do all of the problems. You will have as much time as you need."

SCORING SPECIFICATIONS

All items will be scored dichotomously (0 for incorrect response, 1 for correct response).

REPLACEMENT SCHEME

Use any of the problems from the replacement sets. Any letter of the alphabet may be substitued for another, except for X. Choose one from each of the following sets.

Replacement Sets

Set 1:
$$X(N + R) = Y$$
$$(N + R)X = Y$$

Set 2:
$$\frac{X}{R} + N = Y$$
$$N + \frac{X}{R} = Y$$

Set 3:
$$\frac{AX}{B} + C = D$$
$$C + \frac{AX}{B} = D$$

ITEM FORM 3

Solving Algebraic Problems involving <u>Factoring</u>

GENERAL DESCRIPTION

The individual is given an algebraic equation involving four to six letters and is asked to solve for an unknown. by isolating that unknown on either side of the equation. In problem, the variable X will appear at least twice, one one side of the equation.

STIMULUS AND RESPONSE CHARACTERISTICS

Constant for all Cells

Algebraic problems are presented in written form. A written response is required.

Distinguishing Among Cells

Type of factoring involved:
1) factoring (two steps); 2) factoring (three steps); 3) factoring (four steps); 4) factoring (five steps).

Varying Within Cells

Letters used to represent variables.

CELL MATRIX

two step	1
three step	2
four step	3
five step	4

figure 3.   Item form 3 — Factoring

figure 3 (continued)

ITEM FORM SHELL

Materials

Pencil
Problem sheet with space for responses

Directions to Experimenter	Script
Place materials in front of individual	"Here is a sheet with problems. Solve for X in each problem. Try to do all of the problems. You will have as much time as you need."

SCORING SPECIFICATIONS

All items will be scored dichotomously (0 for incorrect response,1 for correct response).

REPLACEMEMT SCHEME

Use any of the problems from the replacement sets. Any letter of the alphabet may be substituted for another, except for X. Choose one from each of the following sets.

Replacement Sets

Set 1:

$$NX + RX = Y$$
$$RX + NX = Y$$

Set 2:

$$(NX + RX)Y = Z$$
$$Y(NX + RX) = Z$$

Set 3:

$$\frac{Z(NX + RX)}{Y} = P$$
$$\frac{(NX + RX)Z}{Y} = P$$

Set 4:

$$\frac{AX + BX}{D} + C = E$$
$$C + \frac{AX + BX}{D} = E$$

ITEM FORM 4

Solving Algebraic Problems involving the Distributive Property

GENERAL DESCRIPTION

The individual is given an algebraic problem involving four to six letters and is asked to solve for an unknown, X, by isolating that unknown on either side of the equation. For each problem, it will be necessary for the individual to distribute an expression, that is multiplied by another variable.

STIMULUS AND RESPONSE CHARACTERISTICS

Constant for all Cells

Algebraic problems are presented in written form. A written response is required.

Distinguishing among Cells

Number of steps necessary to solve the problem. 1) three steps; 2) four steps; 3) five steps.

Varying within Cells

Letters used to represent variables.

CELL MATRIX

three steps	1
four steps	2
five steps	3

figure 4. Item form 4 - Distributive Property

figure 4 (continued)

ITEM FORM SHELL

Materials

Pencil
Problem sheet with space for response

Directions to Experimenter	Script
Place materials in front of individual	"Here is a sheet with problems. Solve for X in each problem. Try to do all of the problems. You will have as much time as you need."

SCORING SPECIFICATIONS

All items will be scored dichotomously (0 for incorrect response, 1 for correct response).

REPLACEMENT SCHEME

Use any of the problems from the replacement sets. Any letter of the alphabet may be substituted for another, except for X. Choose one from each of the following sets.

Replacement Sets

Set 1:

$$A(X + B) = D$$
$$(X + B)A = D$$

Set 2:

$$\frac{N(X + R)}{Y} = Z$$
$$\frac{(X + R)N}{Y} = Z$$

Set 3:

$$\frac{N(X + R)}{Y} + Z = P$$
$$Z + \frac{N(X + R)}{Y} = P$$

ITEM FORM 5

   Solving Algebraic Problems involving <u>Factoring</u> and the <u>Distributive</u> <u>Property</u>

<u>GENERAL DESCRIPTION</u>

   The individual is given an algebraic equation involving four to six letters and is asked to solve for an unknown, X, by isolating that unknown on either side of the equation. In each problem it will be necessary to factor and/or distribute an expression.

<u>STIMULUS AND RESPONSE CHARACTERISTICS</u>

<u>Constant for all Cells</u>

   Algebraice problems are presented in written form. A written response is required

<u>Distinguishing among Cells</u>

   Number of steps necessary to solve the problem. 1) two steps; 2) three steps; 3) four steps; 4) five steps.

<u>Varying within Cells</u>

   Letters used to represent variables.

<u>CELL MATRIX</u>

two step	1
three step	2
four step	3
five step	4

figure 5. Item form 5 - Factoring and the Distributive Property

figure 5 (continued)

ITEM FORM SHELL

Materials

Pencil
Problem sheet with space for response

Directions to Experimenter	Script
Place materials in front of individual	"Here is a sheet with problems. Solve for X in each problem. Try to do all of the problems You will have as much time as you need."

SCORING SPECIFICATIONS

All items will be scored dichotomously (0 for incorrect response, 1 for correct response).

REPLACEMENT SCHEME

Use any of the problems from the replacement sets. Any letter of the alphabet may be substituted for another, except for X. Choose one from each of the following sets.

Replacement Sets

Set 1:
$$NX + RX = Y$$
$$RX + NX = Y$$

Set 2:
$$(NX + RX)Y = Z$$
$$Y(NX + RX) = Z$$
$$A(X + B) = D$$
$$(X + B)A = D$$

Set 3:
$$\frac{Z(NX + RX)}{Y} = P$$
$$\frac{(NX + RX)Z}{Y} = P$$
$$\frac{N(X + R)}{Y} = Z$$
$$\frac{(X + R)N}{Y} = Z$$

Set 4:
$$\frac{AX + BX}{D} + C = E$$
$$C + \frac{AX + BX}{D} = E$$
$$\frac{N(X + R)}{Y} + Z = P$$
$$Z + \frac{N(X + R)}{Y} = P$$

Appendix B

Algebra Test

Solve for X in each of the following equations.

$$\frac{X}{A} = B$$

$$X(N + R) = Y$$

$$X + (A + B) = C$$

$$\frac{NX}{R} = Y$$

$$\frac{AX}{B} + C = D$$

$$X + A = B$$

$$\frac{AX + BX}{D} + C = E$$

$$\frac{N(X + R)}{Y} + Z = P$$

$$\frac{Z(NX + RX)}{Y} = P$$

$$\frac{X}{R} + N = Y$$

$$A(X + B) = D$$

$$NX + RX = Y$$

$$(NX + RX)Y = Z$$

$$\frac{N(X + R)}{Y} = Z$$

$$X(A + B) = C$$

$$\frac{X}{N} = R$$

$$X + (N + R) = Y$$

$$\frac{AX}{B} = C$$

$$\frac{NX}{R} + Y = Z$$

$$X + N = R$$

$$\frac{NX + RX}{Z} + Y = P$$

$$\frac{A(X + B)}{C} + D = E$$

$$\frac{D(AX + BX)}{C} = E$$

$$\frac{X}{B} + A = C$$

$$N(X + R) = Z$$

$$AX + BX = C$$

$$(AX + BX)\ C = D \qquad\qquad \frac{A(X + B)}{C} = D$$

Appendix C

Tables

Table 1

Observed Cross-Classifications for the Term-Transposition Domain[1]

Response Patterns		Cross-Classifications					
Tasks							
A	B	A1 – M1	A1 – A2	A1 – M2	M1 – A2	M1 – M2	A2 – M2
1	1	65	72	69	82	99	97
1	2	4	2	2	12	2	22
1	3	6	1	4	9	2	16
2	1	14	22	28	14	16	6
2	2	12	12	4	13	10	10
2	3	12	4	6	8	9	10
3	1	24	19	38	17	20	10
3	2	19	40	20	29	14	22
3	3	161	145	146	133	145	124

1. The letters in the letter-number combinations labeling the columns below
    the cross-classifications heading indicate addition-subtraction (A) or
    multiplication-divison (M) problems. The numbers refer to the number of
    steps required for problem solution.

Table 2

Observed Cross-Classifications for the Expressions-Transposition Domain[1]

Response Patterns		Cross-Classifications		
Tasks				
A	B	E1 – E2	E1 – E3	E2 – E3
1	1	135	133	158
1	2	2	11	8
1	3	15	8	0
2	1	10	12	2
2	2	3	4	6
2	3	12	9	15
3	1	21	23	8
3	2	18	20	21
3	3	101	97	99

1. The letter-number combinations labeling the columns indicate the expressions domain and the number of steps required for problem solution. For example E1 indicates an expressions problem in which one step is required for solution.

Table 3

Observed Cross-Classifications for the Factoring-Distributive Property Domain[1]

Cross-Classifications

Response Patterns

Tasks		F2-F3	F2-F4	F2-F5	F2-D3	F2-D4	F2-D5	F3-F4	F3-F5	F3-D3	F3-D4	F3-D5	F4-F5	F4-D3	F4-D4	F4-D5	F5-D3	F5-D4	F5-D5	D3-D4	D3-D5	D4-D5
A	B																					
1	1	160	157	159	144	151	154	169	169	150	161	166	175	156	168	175	159	172	185	152	161	162
1	2	2	6	1	11	18	16	11	2	8	11	12	1	3	6	10	3	4	3	17	8	6
1	3	1	0	3	15	25	41	0	11	12	22	33	7	11	20	26	8	18	23	1	1	0
2	1	11	13	15	10	8	5	9	14	12	8	6	15	16	9	8	19	12	13	8	19	26
2	2	10	6	5	9	5	6	6	2	5	3	5	3	4	3	5	1	1	5	15	4	13
2	3	10	12	11	15	9	32	14	13	17	11	32	14	14	10	30	14	9	25	13	11	17
3	1	9	13	28	9	4	4	5	21	14	7	4	12	19	14	8	24	18	4	8	31	23
3	2	17	20	5	11	8	9	15	7	6	5	2	7	14	12	6	7	6	3	26	31	24
3	3	97	90	90	93	89	50	88	80	93	89	54	83	80	75	49	82	77	56	79	51	46

1. Letters in the cross-classification columns indicate factoring (F) and distributive property (D) problems. Numbers refer to the number of steps required for problem solution.

Table 4

Chi-Squared Tests for the Term-Transposition Domain[1]

Tasks		H_o			H_1			H_2			H_2'			H_3			H_3'			H_4		
A	B	X^2	d.f.	p	X^2	d.f.	p	X^2	d.f.	p	X^2	d.f.	p	X^2	d.f.	p	X^2	d.f.	p	X^2	d.f.	p
A1-M1		200.65	5	<.01	19.04	3	<.01	1.18*	3	<.90	.29	2	<.90	1.19	2	<.50	.16	1	<.90	.87	2	<.75
A1-A2		251.78	5	<.01	73.48	3	<.01	10.92	3	<.025	9.71	2	<.01	10.94	2	<.01	1.38*	1	<.25	7.73	2	<.025
A1-M2		237.56	5	<.01	66.65	3	<.01	29.12	3	<.01	6.73	2	<.05	29.14	2	<.01	2.63*	1	<.25	5.75	2	<.10
M1-A2		197.55	5	<.01	15.32	3	<.01	3.36*	3	<.50	1.93	2	<.50	3.37	2	<.25	1.35	1	<.25	3.27	2	<.25
M1-M2		328.54	5	<.01	32.58	3	<.01	2.58*	3	<.50	.06	2	<.175	2.59	2	<.50	.04	1	<.90	1.07	2	<.75
A2-M2		203.97	5	<.01	15.73	3	<.01	6.69	3	<.10	.41*	2	<.90	6.50	2	<.05	0	1	<.99	6.70	2	<.05

1. The letters in the letter-number combinations used in designating task pairs indicate the addition-subtraction (A) and multiplication-division (M) dimensions. Numbers refer to the number of steps required to achieve problem solution.

Table 5

Chi-Squared Tests for the Expressions-Transposition Domain[1]

Tasks		H_o			H_1			H_2			H_2'			H_3			H_3'			H_4		
A	B	X^2	d.f.	p	X^2	d.f.	p	X^2	d.f.	p	X^2	d.f.	p	X^2	d.f.	p	X^2	d.f.	p	X^2	d.f.	p
E1-E2		302.59	5	<.01	8.03	3	<.025	7.47*	3	<.10	4.44	2	<.25	8.98	2	<.02	3.83	1	<.10	4.04	2	<.25
E1-E3		272.38	5	<.01	11.89	3	<.01	4.88*	3	<.25	3.33	2	<.25	4.88	2	<.10	.06	1	<.90	2.53	2	
E2-E3		421.55	5	<.01	15.95	3	<.01	27.43	3	<.01	15.66	2	<.01	27.44	2	<.01	11.40	1	<.01	27.54	2	<.01

1. Letter-number combinations designating task pairs refer to expressions problems (E) and indicate the number of steps required for problem solution.

Table 6

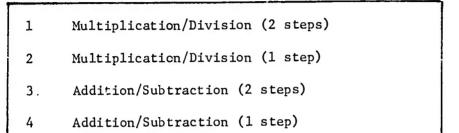Chi-Squared Tests for the Factoring-Distributive Property Domain

Tasks A–B	H_o X^2	d.f.	p	$\bar{n}_1$ X^2	d.f.	p	H_2 X^2	d.f.	p	H_2' X^2	d.f.	p	H_3 X^2	d.f.	p	H_3' X^2	d.f.	p	H_4 X^2	d.f.	p
D3–F2	305.52	5	<.01	1.6086*	3	<.75	.33	3	<.975	.28	2	<.90	.34	2	<.90	.02	1	<.90	.32	2	<.90
D3–F3	345.09	5	<.01	6.4418*	3	<.10	6.51	3	<.10	1.66	2	<.50	6.51	2	<.05	1.07	1	<.50	4.76	2	<.10
D3–F4	337.67	5	<.01	11.924	3	<.01	10.35	3	<.025	3.52	2	<.25	10.39	2	<.01	1.99	1	<.25	.76*	2	<.75
D3–F5	382.18	5	<.01	29.4	3	<.01	30.40	3	<.01	3.63	2	<.25	31.02	2	<.25	0	1	<1	11.97	2	<.01
D3–D4	323.14	5	<.01	13.93	3	<.01	25.33	3	<.01	11.71	2	<.01	25.34	2	<.01	4.27*	1	<.05	25.34	2	<.01
D3–D5	285.98	5	<.01	49.99	3	<.01	15.28	3	<.01	13.55	2	<.01	15.35	2	<.01	2.25*	1	<.25	9.60	2	<.01
D4–F2	292.64	5	<.01	20.939	3	<.01	2.06*	3	<.75	2.02	2	<.50	2.07	2	<.50	1.18	1	<.50	1.53	2	<.25
D4–F3	340.53	5	<.01	0.929	3	<.01	2.69*	3	<.50	.86	2	<.75	2.69	2	<.50	0	1	<1	1.03	2	<.75
D4–F4	320.66	5	<.01	1.85072*	3	<.01	4.54*	3	<.25	2.08	2	<.50	4.56	2	<.25	1.55	1	<.25	.15	2	<.95
D4–F5	363.86	5	<.01	4.79*	3	<.01	20.33	3	<.01	1.92	2	<.50	20.88	2	<.01	.67	1	<.50	6.66	2	<.05
D4–D5	325.15	5	<.01	27.57	3	<.01	12.39	3	<.01	11.21	2	<.01	12.43	2	<.01	.86*	1	<.50	11.41	2	<.01
D5–F2	234.24	5	<.01	55.13	3	<.01	5.32*	3	<.25	2.10	2	<.50	5.32	2	<.10	.02	1	<.90	5.32	2	<.10
D5–F3	310.60	5	<.01	59.904	3	<.01	3.14*	3	<.50	3.14	2	<.25	3.14	2	<.25	1.83	1	<.25	3.15	2	<.25
D5–F4	305.29	5	<.01	12.4176	3	<.01	2.29*	3	<.75	2.06	2	<.50	2.29	2	<.50	.73	1	<.50	2.26	2	<.50
D5–F5	387.10	5	<.01	41.264	3	<.01	6.35*	3	<.10	1.12	2	<.75	6.35	2	<.05	1.12	1	<.50	.45	2	<.90

1. The letters in the task descriptions refer to factoring (F) and distributive property (D) problems. The numbers indicate number of steps to problem solution.

Table 6 (continued)

A B	H_o X^2	d.f.	p	H_1 X^2	d.f.	p	H_2 X^2	d.f.	p	H_2' X^2	d.f.	p	H_3 X^2	d.f.	p	H_3' X^2	d.f.	p	H_4 X^2	d.f.	p
F2-F3	434.24	5	<.01	16.06	3	<.01	2.81*	3	< .50	1.66	2	<.50	2.82	2	<.25	0	1	1	2.80	2	<.25
F2-F4	392.54	5	<.01	22.68	3	<.01	12.63	3	<.01	12.50	2	<.01	12.63	2	<.01	4.58*	1	<.05	11.85	2	<.01
F2-F5	420.47	5	<.01	40.26	3	<.01	23.67	3	<.01	.77	2	<.75	23.68	2	<.01	.06	1	<.90	17.24	2	<.01
F3-F4	427.12	5	<.01	7.17*	3	<.01	23.69	3	<.01	20.09	2	<.01	23.70	2	<.01	5.92*	1	<.025	23.70	2	<.01
F3-F5	384.90	5	<.01	15.13	3	<.01	19.84	3	<.01	1.17	2	<.75	20.18	2	<.01	.52	1	<.50	7.38	2	<.025
F4-F5	441.56	5	<.01	18.41	3	<.01	12.40	3	<.01	2.63	2	<.50	12.42	2	<.01	1.72	1	<.25	1.44*	2	<.50

Table 7

Empirically Validated Equivalence Relations for Three

Homogeneous Item Domains[1]

Domain 1 - Term Transposition

1	Multiplication/Division (2 steps)
2	Multiplication/Division (1 step)
3.	Addition/Subtraction (2 steps)
4	Addition/Subtraction (1 step)

Domain 2 - Expression Transposition

1	Three Steps
2	Two Steps
3	One Step

Domain 3 - Factoring/Distributive Property

1	Distributive (4 steps)/Factoring (4 steps)
2	Distributive (3 steps)/Factoring (3 steps)
3	Distributive (3 steps)/Factoring (2 steps)

1. Numbers represent difficulty levels within a domain. The numeral 1 represents the most difficult, the numeral 2 represents the next to the most difficult, etc.

Appendix D

ALT Computer Program

A Computer Program for Multidimensional Contingency

Table Construction[1]

John R. Bergan   Olga M. Towstopiat

John W. Luiten

The University of Arizona

Abstract

A Computer Program for Multidimensional Contingency

Table Construction[1]

John R. Bergan   Olga M. Towstopiat

John W. Luiten

The University of Arizona

A computer program (ALT) is presented as an aid in the analysis
of cross-classified categorical data. Program ALT can accept a
maximum of 500 variables as data and construct a frequency table
of from 1 to 8 dimensions. Each variable may have as many as 8
categories. In addition, ALT can combine, recode, and create for-
mal variables at the user's discretion.

A Computer Program for Multidimensional Contingency Table Construction

John R. Bergan   Olga M. Towstopiat

John W. Luiten

The University of Arizona

During the past 15 years, many new developments have occurred in the analysis of cross-classified categorical data. The development of loglinear models (Bishop, Feinberg, and Holland, 1975; Goodman, 1972) has made it possible to assess effects among dichotomous and polytomous categorical variables. The establishment of quasi-independence techniques (Goodman, 1968, 1975) has been useful in the analysis of occupational mobility tables and in response scaling. Advances in latent structure analysis have also made it possible to identify restricted and unrestricted latent class models representing relations involving categorical variables (Goodman, 1974a, 1974b).

In many cases, researchers analyzing cross-classified categorical data are faced with the task of selecting a subset of variables from a larger set for use in contingency table construction. For example, a researcher may administer a questionnaire with 50 dichotomous items and later wish to analyze relationships among several subsets of those items. When the number of subjects is large, it can be a formidable task to tally the frequencies needed to construct contingency tables to examine relationships among subsets of variables. The purpose of this paper is to describe a computer program for use in multidimensional contingency table construction.

The Program

The program described in this article enables investigators to select a subset of variables from a larger set, to construct a contingency table from those variables, and to count the frequency of responses in the various cells of the table. Moreover, this program can combine and recode variables in table construction. In addition, it can be attached to other widely-used computer programs for the analysis of cross-classified data. For instance, the program can be linked to Fay and Goodman's (1973) ECTA (Everyman's Contingency Table Analyzer). It may also be attached to Clifford Clogg's (1977) MLLSA (Maximum Likelihood Latent Structure Analysis) Program.

Input

Program ALT is written in Fortran IV Extended for use on the Cyber 175 computer. A maximum number of 500 variables may be entered as data for processing by the program. The program will then select from 1 to 8 of these variables and compute the cell frequencies for a contingency table formed from the selected variables. For example, if a researcher had a questionnaire with 100 items requesting yes or no responses, he/she could choose to have the responses to four of the items tabulated into a four-dimensional contingency table.

The variables may be categorized as dichotomies or as polytomies, with a maximum of 8 categories. Thus, the variables selected could contain pass/fail scores or responses scaled from 1 to 8. The program allows a maximum of 300 cells per table, regardless of whether dichotomous or polytomous categories are employed.

Any type of data may be entered into the program, including integers and real numbers with decimal points. However, if polytomous categories are not comprised of consecutive values, the recode subroutine must be used to recode the data. For example, if an investigator had a series of IQ scores within the 80-89, 100-109, and 110-119 ranges, the scores between 80 and 89 could be recoded as 1, the scores between 100 and 109 as 2, and the scores between 110 and 119 as 3.

Data Altering

There are three data altering subroutines in the program. One is used to combine variables (Subroutine 1); the second recodes variables (Subroutine 2); and the third is employed when the user wishes to construct a contingency table involving formal variables (Subroutine 3) (Duncan, 1975). The user can call any of the three subroutines independently. However, the subroutines are often employed together. For example, a user might wish to apply the combined variables subroutine to construct a new variable formed by combining two or more of the initially selected variables. The user might then wish to recode the scores for the newly constructed variable.

Subroutine 1. The combined variables subroutine enables the user to select from two or more variables to be combined in an additive fashion. For example, a user could combine three variables into one. If this procedure were followed, the scores for each of the three variables would be added together to yield a score for each subject on the newly-constructed variable. If desired, the

program can print scores for each subject for newly constructed variables obtained after application of the combined variables subroutine.

Use of the combined variables subroutine makes it possible to construct contingency tables composed entirely of variables resulting from combinations, or tables in which some of the variables have been combined and others have not been combined. The program can print the total number of variables that have been combined as well as the specific variables that have been used in each combination.

Subroutine 2. The recode subroutine enables the user to recode variable scores for each subject. For example, if a user wishes to describe a variable such as a selected range of mathematics achievement scores in terms of several categories, mathematics scores from 70-79 could be recoded as 1, scores from 80-89 recoded as 2, scores from 90-99 as 3, etc. If desired, the program will print recoded scores for each subject. In addition, printed output specifies the minimum and maximum values used as a basis for establishing recoded categories. For example, if the minimum value for category 1 were 85 and the maximum value were 92, the program would print the 85 and 92 for category 1.

Subroutine 3. When a researcher is working with a polytomous variable, it is generally useful to be able to identify the contribution of specific categories within the polytomy to association in the table under examination. One way to accomplish this task is to re-express the polytomous variable as a set of formal variables, each indicating the contrast between one category of the polytomy

and all other categories (Duncan, 1975). For example, suppose that the polytomous category Protestant, Catholic, and Jewish were associated with a yes-no dichotomy in a two-way contingency table. It would probably be of interest to determine the contribution to association in the table of each of the three categories in the polytomy. Re-expressing the polytomy as a set of formal variables could be a first step toward determining the contribution to association of the specific categories in the polytomy.

The formal variables subroutine enables the user to construct a table in which a polytomy can be re-expressed in terms of a set of formal variables. Table 1 shows the construction of an 8 x 2 contingency table. The original table is shown under A and the altered table is shown under B.

Insert Table 1 about here

Output

The descriptive information provided by the program is given in this section. Titles of the output are provided. In addition, descriptive statements regarding the output are included in cases in which the output titles are not entirely self-explanatory.

1. Minimum and maximum values for variable categories. The upper and lower numerical limits for each variable category are printed.

2. Title of user's data run.

3. Number of subjects.

4. Number of variables per subject.

5. <u>Number of variables selected</u>. The total number of variables selected by the user for the construction of a contingency table is printed.

6. <u>Print/punch option chosen</u>. The program prints an <u>0</u> if the program output was printed and no cards were punched, and a <u>1</u> if the program output was printed and cards were punched.

7. <u>Number of combined variables</u>. The number of combined variables obtained after employing the combined variables subroutine is printed.

8. <u>Variable combinations</u>. The sets of variables that were combined are printed.

9. <u>Selected variables</u>. The numbers assigned to the variables chosen for the construction of the contingency table are printed.

10. <u>Response matrix</u>. The frequencies for each of the cells in the contingency table are printed.

11. <u>Subject scores with combined variables</u>. The final combined variable scores for each subject may be printed.

12. <u>Recoded scores for each subject</u>. The recoded variable scores for each subject may be printed. If the variable scores are processed by the combined variables subroutine, the resultant combined scores will be recoded and printed.

Limitation

A technical limitation to the program's usefulness is the restriction of selecting a maximum of 8 dimensions per table. This restriction may be eliminated if the user alters the dimension statements within the program.

Summary

Program ALT selects a subset of variables from a larger set, constructs a contingency table from the variable subset, and then calculates cell frequencies for the contingency table. The resultant cell frequencies may then be used to conduct further statistical analyses with programs such as the MLLSA or ECTA.

The program's flexibility allows the user to have (1) the data read from disc files, tape files, or computer cards; (2) the variables entered in Fortran order with the variable on the far left varying most rapidly ($.g., X_{111}, X_{211}, etc....) or with the variable on the far right changing most often; (3) a virtually unlimited number of subjects' scores tabulated (i.e., 99,999); and (4) have printed and/or punched card output.

Availability

Individuals interested in obtaining a program listing and specific instructions for program use should write to the following address:

John Bergan

Department of Educational Psychology

College of Education

The University of Arizona

Tucson, Arizona 85721

References

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. Discrete multivariate analysis. Cambridge, Mass.: MIT Press, 1975.

Clogg, C.C. Unrestricted and restricted maximum likelihood latent structure analysis: A manual for users. Unpublished document available from C.C. Clogg, Pennsylvania State University, 1977.

Duncan, O.D. Partitioning polytomous variables in multiway contingency analysis. Social Science Research, 1975, 4, 167-182.

Fay, R., and Goodman, L.A. ECTA Program description for users. Unpublished document available from L.A. Goodman, University of Chicago, 1973.

Goodman, L.A. The analysis of cross-classified data: Independence, quasi-independence, and interaction in contingency tables with or without missing cells. Journal of the American Statistical Association, 1968, 63, 1091-1131.

Goodman, L.A. A general model for the analysis of surveys. American Journal of Sociology, 1972, 77, 1035-1085.

Goodman, L.A. The analysis of systems of quantitative variables when some of the variables are unobservable. Part I: A modified latent structure approach. American Journal of Sociology, 1974a, 79, 1179-1259.

Goodman, L.A. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika, 1974b, 61, 2, 215-230.

Goodman, L.A. A new model for scaling response patterns: an application of the quasi-independence concept. Journal of the American Statistical Association, 1975, 70, 755-768.

Footnotes

[1]The development of this program was supported in part by a contract from the Defense Advanced Research Project Agency.

Table 1

Religion by Response Contingency Table[1]

A

	Yes	No
Protestant	10	10
Catholic	30	30
Jewish	10	60

B

Protestant	Catholic	Jewish	Yes 1	No 2
1	1	1	0	0
1	1	2	0	0
1	2	1	0	0
1	2	2	10	10
2	1	1	0	0
2	1	2	30	30
2	2	1	10	60
2	2	2	0	0

[1]A 1 refers to a positive response for assignment to one of the three religious categories, and a 2 refers to a negative response for category assignment.